

A coherent method for the detection and estimation of continuous gravitational wave signals using a pulsar timing array

Yan Wang^{1,2}, Soumya D. Mohanty^{1,3} and Fredrick A. Jenet^{1,2,3}

Department of Physics and Astronomy, University of Texas at Brownsville, TX 78520

Center for Advanced Radio Astronomy, Brownsville, TX 78520

Center for Gravitational Waves Astronomy, Brownsville, TX 78520

ABSTRACT

The use of a high precision pulsar timing array is a promising approach to detecting gravitational waves in the very low frequency regime ($10^{-6} - 10^{-9}$ Hz) that is complementary to the ground-based efforts (e.g., LIGO, Virgo) at high frequencies ($\sim 10 - 10^3$ Hz) and space-based ones (e.g., LISA) at low frequencies ($10^{-4} - 10^{-1}$ Hz). One of the target sources for pulsar timing arrays are individual supermassive black hole binaries that are expected to form in galactic mergers. In this paper, a likelihood based method for detection and estimation is presented for a monochromatic continuous gravitational wave signal emitted by such a source. The so-called pulsar terms in the signal that arise due to the breakdown of the long-wavelength approximation are explicitly taken into account in this method. In addition, the method accounts for equality and inequality constraints involved in the semi-analytical maximization of the likelihood over a subset of the parameters. The remaining parameters are maximized over numerically using Particle Swarm Optimization. Thus, the method presented here solves the monochromatic continuous wave detection and estimation problem without invoking some of the approximations that have been used in earlier studies.

Subject headings: pulsar timing array: general — continuous gravitational waves: detection algorithm

1. Introduction

Several worldwide projects are in progress to open the gravitational wave window for observational astronomy (Sathyaprakash & Schutz 2009). Taken together, these projects will span a wide range of astrophysically promising source frequencies, providing complementary

views of the GW Universe. In the 10 Hz to 1 kHz range, the LIGO (Abbott et al. 2009) and Virgo (Accadia et al. 2011) projects have already performed several joint observational runs (Abadie et al. 2012a,b). Work is now in progress to commission second generation ground based detectors (Advanced LIGO (Waldman 2011), Advanced Virgo (Degallaix et al. 2013), and KAGRA (Somiya 2012)). There have been long-standing plans for the 0.1 mHz to 0.1 Hz regime using space-based detectors. The LISA Pathfinder is scheduled for a launch in 2015. It will demonstrate and test the technologies to be used in the eLISA mission (Seoane et al. 2013). In the very low frequency regime of 10^{-9} to 10^{-6} Hz, the use of Pulsar Timing Arrays (PTA) is currently being explored intensively.

In the past several decades, pulsar timing has produced many significant discoveries in astronomy which include the first evidence of the existence of gravitational waves (Taylor & Weisberg 1982, 1989). In pulsar timing (Lorimer & Kramer 2004), time of arrivals (TOAs) of radio pulses from a rotating pulsar are fitted by a linearized timing model (Edwards et al. 2006). The parameters of the model can be classified into astronomical (e.g. sky location, proper motion, period, period derivatives), interstellar medium (dispersion measure), binary system and instrumental parameters. The differences between the actual and the best-fit arrival times are called timing residuals. By subtracting out all known effects, the timing residuals from a given pulsar should just reduce to pure noise. Any deviation from this expectation may be attributed to the effect of a gravitational wave passing between the pulsar and Earth.

The sensitivity of GW detection using pulsar timing is characterized by the root mean square (RMS) of the timing residuals. Currently a handful of millisecond pulsars (MSPs) have archived RMS at 100 ns level over many years (Verbiest et al. 2009; Demorest et al. 2013). It has been shown that an array of pulsars timed to this level of precision can be operated as a galactic scale instrument to detect very low frequency GWs (Hellings & Downs 1983; Foster & Backer 1990). A detection using such a PTA can be made by observing 20–40 pulsars over 5–10 years, assuming a monthly observation cadence and 100 ns timing precision for each (Jenet et al. 2005b).

At present, there are three major PTAs in operation: the North American Nanohertz Observatory for Gravitational Waves (NANOGrav, Demorest et al. (2013)), the Parkes Pulsar Timing Array (PPTA, Manchester et al. (2013)), and the European Pulsar Timing Array (EPTA, Ferdman et al. (2010)). As independent consortia, the three PTAs compose the International Pulsar Timing Array (IPTA, Manchester (2013)) with approximately 50 pulsars regularly monitored. Data shared between PTAs can form a longer observation duration and finer observation cadence for a specific pulsar (e.g. PSR J1713+0747, Zhu & et al. (2014)). In addition, combining geographically widely distributed telescopes of all the PTAs increases

sky coverage and reduces certain systematics.

One of the principle signals anticipated in PTA based GW detection is the stochastic background formed by the incoherent superposition of GW signals from a large population of unresolved sources distributed throughout the Universe. Besides this background, it may be possible to detect and characterize individual sources using a PTA (Sesana & Vecchio 2010). Using the data from the Millennium Simulation of structure formation (Springel et al. 2005), and considering a broad range of population models for Super Massive Black Hole Binaries (SMBHBs), Sesana et al. (2009) found that sufficiently strong GW signals generated by close and/or massive SMBHBs could stand above the stochastic background and be resolved individually. Pulsar timing has already been used to rule out the radio galaxy 3C 66B (Sudou et al. 2003) as SMBHB candidate due to the absence of any detectable effect on pulsar timing residuals by the GW signal that should have been produced by a binary source (Jenet et al. 2004). Besides SMBHBs, intermediate-mass black hole binaries that may reside in globular clusters could also be potential GW sources for PTA (Jenet et al. 2005a).

Along with observational advances in pulsar timing precision, there are data analysis issues that must be satisfactorily resolved in order to increase the sensitivity of PTA-based GW detection. PTA data presents several features, such as irregular sampling and non-stationary noise, that data analysis algorithms must contend with. Since different algorithms may react differently to these features, it is important that a variety of algorithms be developed and compared. Several different data analysis methods have been proposed for the case of continuous GW signals (Yardley et al. 2010; Corbin & Cornish 2010; Babak & Sesana 2012; Ellis et al. 2012; Ellis 2013).

It is natural that methods for continuous source detection using PTA share much in common with those already in use for ground-based detectors. Following the Generalized Likelihood Ratio Test (GLRT) approach (Kay 1998), Jaranowski et al. (1998) developed the \mathcal{F} -statistic method for a single detector with a known noise model. This method was later generalized by Cutler & Schutz (2005) to the case of a network of detectors with time-varying noise. In the PTA regime, the long wavelength approximation that is used to obtain the \mathcal{F} -statistic is no longer valid. This results in the phase of the gravitational wave at a pulsar being substantially different from the one at Earth. Hence, the timing residual signal due to a GW source acquires additional parameters related to the distance from the pulsar to Earth and the direction of the source relative to the line of sight to the pulsar. There are as many of these additional so-called pulsar phase parameters brought into the maximization of the likelihood as there are pulsars in a PTA.

In the case of the \mathcal{F} -statistic, the likelihood is analytically maximized over a subset of signal parameters. These parameters are called extrinsic, while the ones left over are

called intrinsic. The division of the signal parameters into these subsets is quite natural and unique. However, once the pulsar phase parameters appear in the picture, this division is no longer unique. One option is to choose the same division as is done for the \mathcal{F} -statistic. In this case, the pulsar phases have to be maximized numerically. The other option is to treat the pulsar phases as extrinsic, but this pushes all other parameters into the class of intrinsic parameters. The latter option may well be the direction that needs to be followed, especially when it comes to a PTA with a large number of pulsars or the simultaneous detection and estimation of multiple sources, but it has not been well explored yet. The former option is better understood, but it increases the computational cost of maximizing the likelihood, due to the large number of additional parameters, by orders of magnitude.

Previous analyses of PTA sensitivity to individual SMBHB signals have ignored the pulsar phase parameters (Babak & Sesana 2012), or handled them suboptimally (Ellis et al. 2012). However, Corbin & Cornish (2010) showed that excluding these parameters can drastically reduce sensitivity as well as increase parameter estimation errors. They derive this conclusion from a Bayesian approach where a single data realization was used to derive statistical measures of error from the posterior degree of belief (Gregory 2010). It was also found that an estimation of the pulsar phase parameters could lead to reasonably accurate (compared with parallax measurements derived) estimates of the distances to some of the pulsars in a PTA.

In this paper, we investigate the efficacy of GLRT detection and estimation following the option of treating the pulsar phase parameters as intrinsic. It is shown that the analytic maximization over the extrinsic parameters is actually a constrained optimization problem involving an equality and an inequality constraint, both of which are non-linear. Our proposed method takes these constraints into account explicitly. To overcome the computational barrier posed by the above choice of intrinsic parameters, we use Particle Swarm Optimization (PSO). This optimization method was introduced by Eberhart & Kennedy (1995) and has become quite popular across a wide range of fields. PSO was introduced in GW data analysis by Wang & Mohanty (2010). Its use in PTA has been explored by Taylor et al. (2012), although it was only employed for a two-dimensional search space. The method presented here uses PSO over a 12-dimensional search space.

The performance of our method is investigated using three test cases spanning a wide range of signal strengths from the very strong to the barely detectable. (Quantified using the network signal to noise ratio defined in Sec. 5, the values used are 100, 10 and 5.) We use a large number of independent data realizations and derive conventional Frequentist error estimates for the signal parameters. The strong signal case shows that the estimation of the signal waveform in the timing residual of each pulsar has a weak dependence on the

estimation accuracy of pulsar phase parameters. This suggests that the pulsar phases are actually nuisance parameters and that the likelihood is highly degenerate when considered as a function over these parameters. As the signal strength is reduced to more realistic levels, the estimated pulsar phase parameters acquire a nearly uniform distribution across their allowed range. Therefore, this suggests strongly that in future algorithms, they should be treated as extrinsic parameters (or as parameters that are marginalized over). Within the domain of the approximations made in our study, we find that it should be possible to make a confident detection of an SMBHB signal at astrophysically realistic signal strengths. As an example, for an SMBHB of $10^9 M_\odot$ chirp mass with an orbital period of 0.785 years located at 100 Mpc, for which the observed signal has a network matched filtering signal to ratio of 10, the detection probability is ≈ 1 at a false alarm probability $\sim 10^{-3}$. The one-sigma contour of the estimated sky location encloses an area of 800 deg^2 , but the orbital frequency can still be estimated with a standard deviation $< 0.1 \text{ rad yr}^{-1}$.

The rest of the paper is organized as follows. In Sec. 2 we introduce the data model used in this paper. Sec. 3 describes the GLRT for this data model which involves constrained maximization over extrinsic parameters. Sec. 4 presents a brief review of PSO and discusses its implementation in this application. Sec. 5 demonstrates the method and quantifies its performance using test cases. The paper is concluded in Sec. 6.

2. Data model

The data from a PTA consists of a set of sequences of timing residuals, one for each pulsar, $r^I = (r_1^I, r_2^I, \dots, r_{N_I}^I)$, $I = 1, 2, \dots, N_p$, where N_p is the number of pulsars, N_I is the number of observation for the I -th pulsar, and r_i^I is the timing residual observed at time $t_i^I \in [0, T]$, $t_{i+1}^I > t_i^I$, by an observer at Earth.

$$r_k^I = s_k^I + n_k^I \text{ when a GW signal is present,} \quad (1)$$

$$= n_k^I \text{ when a GW signal is absent.} \quad (2)$$

where $n^I = (n_1^I, n_2^I, \dots, n_{N_I}^I)$ and $s^I = (s_1^I, s_2^I, \dots, s_{N_I}^I)$ denote the noise realization and the GW signal in the I -th timing residual sequence. In this section, we first introduce the GW signal family used in this paper, followed by a description of the noise model.

2.1. Timing residuals induced by GW

Consider the scenario shown in Fig. 1 where plane GWs from some distant continuous source, such as a SMBHB, are crossing the line of sight from Earth to a Galactic pulsar. The

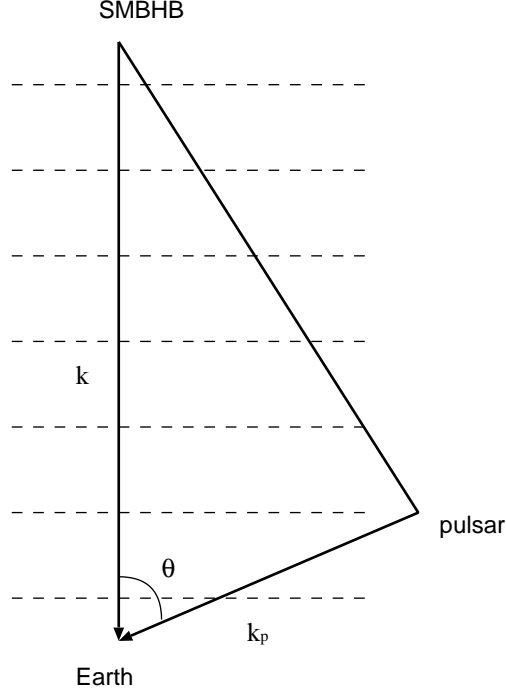


Fig. 1.— Supermassive black hole binary, pulsar and Earth configuration. θ is the open angle between the supermassive black hole binary and the pulsar subtended at Earth. \mathbf{k}_p is the wave vector of a radio pulse, \mathbf{k} is the wave vector of GWs. Dash lines represent the wave front of the GWs.

stretching and squeezing of the path length of radio pulses by the GW cause fluctuations in the arrival time of the pulses at Earth. Even though individual radio pulses from a pulsar are emitted irregularly in time, integrated pulses have a predictable TOA (Lorimer & Kramer 2004). Any deviation from the predicted TOAs is called the timing residual, part of which may be attributed to the effect of a GW. The effect of the GW signal on the timing residual is achromatic, i.e. the magnitude of the fluctuation is independent of the carrier frequency of the radio emission. The fluctuation of a specific pulse can be calculated by integrating the fractional change in frequency of the pulse rate (time redshift) induced by GW along the trajectory from the pulsar to the observer. The GW tensor is

$$\mathbf{h} = (A_+ \mathbf{e}_+ + A_\times \mathbf{e}_\times) e^{i(\omega_{gw} t - \mathbf{k} \cdot \mathbf{x})}, \quad (3)$$

where ω_{gw} is the angular frequency, \mathbf{k} is the wave vector of GW. For a non-evolving binary with circular orbit, which is considered in this paper, $\omega_{gw} = 2\omega$ (ω is the angular frequency of the binary). $\mathbf{e}_+ = \hat{\alpha} \otimes \hat{\alpha} - \hat{\delta} \otimes \hat{\delta}$ and $\mathbf{e}_\times = \hat{\alpha} \otimes \hat{\delta} + \hat{\delta} \otimes \hat{\alpha}$ are the basis tensors of + and \times polarizations, and $\hat{\alpha}$ and $\hat{\delta}$ are the basis vectors of the right ascension and declination of

the equatorial coordinates. Expressions for A_+ and A_\times can be found in Maggiore (2007). Without loss of generalization, it is assumed in the following that the pulsar and observer are both fixed in an inertial reference frame (their motions relative to this reference frame can be treated separately as an additive effect). For I -th pulsar, the timing residual induced by a monochromatic continuous GW signal for an observer at Earth is

$$\begin{aligned} s_k^I &= \int_0^{t_k^I} z(t') dt' = s_e^I(t_k^I) - s_p^I(t_k^I), \\ &= \Re \left\{ \frac{1}{2i} \cdot \frac{\mathbf{k}_p^I \cdot \mathbf{A} \cdot \mathbf{k}_p^I}{1 - \cos \theta^I} \cdot e^{i(2\varphi_0 + \omega_{gw} t_k^I)} \left(1 - e^{-i\omega_{gw} d_p^I (1 - \cos \theta^I)} \right) \right\}. \end{aligned} \quad (4)$$

where $z(t')$ is the redshift of the pulse at time t' , s_e^I and s_p^I are the so called Earth term and pulsar term respectively, \Re is the real part of the expression in the parenthesis, \mathbf{k}_p^I is the wave vector of pulse, φ_0 is the initial phase of the binary orbit at the starting time of the observations, θ^I is the open angle between the SMBHB and the pulsar subtended at the Earth, and d_p^I is the distance between the pulsar and Earth.

The signal can be conveniently rewritten as,

$$s_k^I = \sum_{\mu=1}^4 a_\mu A_\mu^I(t_k^I). \quad (5)$$

The coefficient a_μ are given by

$$a_1 = -\zeta(1 + \cos^2 \iota) \cos 2\psi, \quad (6a)$$

$$a_2 = -2\zeta \cos \iota \sin 2\psi, \quad (6b)$$

$$a_3 = \zeta(1 + \cos^2 \iota) \sin 2\psi, \quad (6c)$$

$$a_4 = -2\zeta \cos \iota \cos 2\psi, \quad (6d)$$

where ι is the inclination angle between the binary orbital plane and the plane of the sky, ψ is the GW polarization angle, and

$$\zeta = \frac{G\mu a^2 \omega^2}{c^4 D} = \frac{G^{5/3}}{c^4 D} \mathcal{M}_c^{5/3} \omega^{-1/3}, \quad (7)$$

where μ is the reduced mass, $\mathcal{M}_c = \mu^{3/5} M^{2/5}$ is the chirp mass of the binary source, M is its total mass, a is its semi-major axis, and D is the distance from Earth to the source. These coefficients only contain the source parameters ζ, ι, ψ which are shared among all pulsars.

The basis functions $A_\mu^I(t_k^I)$ are

$$A_1^I(t_k^I) = -2F_+^I \sin(\varphi_0 - \varphi_I) \sin(\varphi_0 + \varphi_I + \Phi(t_k^I)), \quad (8a)$$

$$A_2^I(t_k^I) = 2F_+^I \sin(\varphi_0 - \varphi_I) \cos(\varphi_0 + \varphi_I + \Phi(t_k^I)), \quad (8b)$$

$$A_3^I(t_k^I) = -2F_\times^I \sin(\varphi_0 - \varphi_I) \sin(\varphi_0 + \varphi_I + \Phi(t_k^I)), \quad (8c)$$

$$A_4^I(t_k^I) = 2F_\times^I \sin(\varphi_0 - \varphi_I) \cos(\varphi_0 + \varphi_I + \Phi(t_k^I)), \quad (8d)$$

where the parameters are: the right ascension α and declination δ of the source, ω_{gw} , φ_0 , and the pulsar phase φ_I (considering that the right ascension α_p^I and declination δ_p^I of the pulsar are known with high precision.) Note that $\varphi_I = \varphi_0 - \omega d_p^I(1 - \cos \theta^I)$ defined for pulsar I will be regarded as an independent parameter in this analysis. $\Phi(t)$ is the phase evolution of the GWs. For circular orbit, $\Phi(t_k^I) = \omega_{gw} t_k^I$. Here $F_+^I = P_+^I/(1 - \cos \theta^I)$ and $F_\times^I = P_\times^I/(1 - \cos \theta^I)$ are the *antenna pattern functions* with P_+^I and P_\times^I are defined as ($\tilde{\alpha}^I = \alpha - \alpha_p^I$)

$$P_+^I = -\cos^2 \delta_p^I (1 - 2 \cos^2 \tilde{\alpha}^I + \cos^2 \tilde{\alpha}^I \cos^2 \delta) + \sin^2 \delta_p^I \cos^2 \delta - \sin 2\delta_p^I \cos \tilde{\alpha}^I \sin \delta \cos \delta, \quad (9)$$

and

$$P_\times^I = 2 \cos \delta_p^I \sin \tilde{\alpha}^I (-\sin \delta_p^I \cos \delta + \cos \delta_p^I \cos \tilde{\alpha}^I \sin \delta). \quad (10)$$

For a supermassive black hole binary, the overall amplitude is

$$\zeta \approx 5 \times 10^{-7} \left(\frac{\mathcal{M}_c}{10^9 M_\odot} \right)^{5/3} \left(\frac{D}{10 \text{ Mpc}} \right)^{-1} \left(\frac{P}{5 \text{ yr}} \right)^{1/3} \text{ sec}. \quad (11)$$

where $P = 2\pi/\omega$ is the orbital period of the binary.

2.2. The noise model

It is commonly assumed in studies of PTA data analysis methods that the noise process in timing residuals is Gaussian and stationary. The probability density function of the noise time series can be characterized by 5 parameters for each pulsar (Arzoumanian et al. 2014).

One way to account for the noise parameters, as is done for the \mathcal{F} -statistic, is to estimate them independently and use the estimates as fixed values in the detection and estimation of signals. This approach is a good approximation when the signal is weak and has no significant effect on the noise parameter estimates. Another approach is to include the noise parameters along with those of the signal in an overall estimation method. This approach leads to significantly higher computational costs and its use in the weak signal case must be examined with care as it may not lead to substantial improvement in the final results.

In this paper, following previous works (Babak & Sesana 2012; Ellis et al. 2012; Arzoumanian et al. 2014), we assume that the noise processes in the N_p pulsars are mutually independent Gaussian and stationary, and use the \mathcal{F} -statistic approach of using independently estimated noise parameters.

Define the *noise weighted inner product* for two time series x and y

$$\langle x|y \rangle_I = x \Sigma_I^{-1} y^T, \quad (12)$$

where $\Sigma_I = E(n^I n^I)$ is the auto-covariance matrix of the noise process in the pulsar I . Then the joint probability density function of the data in the absence of any GW signal is given by

$$p(\mathbf{n}) = \prod_{I=1}^{N_p} p(n^I) = \prod_{I=1}^{N_p} \frac{1}{(2\pi)^{N_I/2} |\Sigma_I|^{1/2}} \exp \left[-\frac{1}{2} \langle n^I | n^I \rangle_I \right]. \quad (13)$$

Here \mathbf{n} denotes the set $\{n^1, n^2, \dots, n^{N_p}\}$. Each n^I is a row vector representing the noise time series for the I -th pulsar, and its length is determined by the number of observations N_I conducted for this pulsar. We have assumed that n^I are mutually independent between pulsars. $|\Sigma_I|$ is the determinant of the auto-covariance matrix.

3. Maximum likelihood ratio statistic

In hypotheses testing, we formulate two mutually exclusive hypotheses for a pulsar timing array data set:

- \mathcal{H}_0 : $r^I(t) = n^I(t)$, $I = 1, \dots, N_p$, there is no signal in the data.
- \mathcal{H}_λ : $r^I(t) = n^I(t) + s^I(t; \lambda)$, $I = 1, \dots, N_p$, there is a signal characterized by parameter λ in the data.

Here \mathcal{H}_λ is a *composite* hypothesis, and $\lambda = \{\zeta, \iota, \psi, \alpha, \delta, \omega_{gw}, \varphi_0, \varphi_I\}$ ($I = 1, 2, \dots, N_p$). One needs to choose between these two hypotheses based on the observed data. Geometrically, it is equivalent to divide the observation space R^N (N -dimensional real space, $N = \sum_{I=1}^{N_p} N_I$) into two disjoint regions R_0 and R_λ . If the data $\mathbf{r} = (r^1, \dots, r^{N_p}) \in R_\lambda$, \mathcal{H}_λ is chosen; while if the data $\mathbf{r} \in R_0$, \mathcal{H}_0 is chosen. The boundary of the two regions is called the *decision surface* D . Different detection strategies are distinguished by different choices of D . In general, there does not exist a D that is optimal for all values of λ . Only in the special case where λ is completely known (or some trivial extensions thereof) does one get an optimal D . Following the *Neyman-Pearson criterion*, in which D minimizes false dismissal probability for a given

false alarm probability, the optimal D turns out to be an iso-surface of the likelihood ratio,

$$\text{LR}(\mathbf{r}) = \frac{p(\mathbf{r}|H_\lambda)}{p(\mathbf{r}|H_0)}. \quad (14)$$

In the case of an unknown λ , a natural modification is to use iso-surfaces of the GLRT functional of data \mathbf{r} obtained by maximizing the likelihood ratio over λ ,

$$\text{GLRT}(\mathbf{r}) = \max_{\lambda} \frac{p(\mathbf{r}|H_\lambda)}{p(\mathbf{r}|H_0)}. \quad (15)$$

Note that the same decision surface can be obtained by replacing the likelihood ratio by any monotonic function, such as the logarithm, of itself. In the following we use the maximum of $\Lambda(\mathbf{r}) = \ln(\text{LR}(\mathbf{r}))$ over λ as the detection statistic.

3.1. The likelihood ratio

The logarithm of the likelihood ratio of the hypotheses in Eq. 14 is

$$\Lambda(\mathbf{r}) = \ln \frac{p(\mathbf{r}|\mathcal{H}_\lambda)}{p(\mathbf{r}|\mathcal{H}_0)} = \sum_{I=1}^{N_p} \langle r^I | s^I(\lambda) \rangle_I - \sum_{I=1}^{N_p} \frac{1}{2} \langle (s^I(\lambda) | s^I(\lambda)) \rangle_I. \quad (16)$$

The detection statistic is the maximum of $\Lambda(\mathbf{r})$ over the parameter space λ . Although the number of parameters involved in the maximization of $\Lambda(\mathbf{r})$ are large, a subset of them can be maximized efficiently using a semi-analytical approach. However, the choice of this subset is not unique. Following the approach used in the \mathcal{F} -statistic, we could pick ζ, ι, ψ as the subset, which would require a numerical maximization over the pulsar phase φ_I and other parameters in the complementary subset. Alternatively, one could choose the φ_I as the subset leading to numerical maximization over the complementary subset. The merits and demerits of these two alternative approaches need much further study. Here, we simply follow the \mathcal{F} -statistic approach and divide the parameters into a subset $\lambda_e = \{\zeta, \iota, \psi\}$, to be maximized semi-analytically, and a subset $\lambda_i = \{\alpha, \delta, \omega_{gw}, \varphi_0, \varphi_I\}$ to be maximized numerically.

For a network of N_p pulsars, and using Eq. 5 we have from Eq. 16

$$\Lambda(\mathbf{r}) = \sum_{\mu=1}^4 a_\mu \sum_{I=1}^{N_p} \langle r^I | A_\mu^I \rangle_I - \frac{1}{2} \sum_{\mu=1}^4 \sum_{\nu=1}^4 a_\mu a_\nu \sum_{I=1}^{N_p} \langle A_\mu^I | A_\nu^I \rangle_I \quad (17)$$

$$= \mathbf{a} \cdot \mathbf{N} - \frac{1}{2} \mathbf{a} \cdot \mathbf{M} \cdot \mathbf{a} \quad (18)$$

where $N_\mu = \sum_{I=1}^{N_p} \langle r^I | A_\mu^I \rangle_I$ is an 4×1 vector which contains the data and the intrinsic parameters $\{\alpha, \delta, \omega_g, \varphi_0, \varphi_I\}$, $M_{\mu\nu} = \sum_{I=1}^{N_p} \langle A_\mu^I | A_\nu^I \rangle_I$ is an 4×4 matrix that contains the

intrinsic parameters only. The unconstrained maximum (Babak & Sesana 2012; Ellis et al. 2012) of $\Lambda(\mathbf{r})$ is easily obtained from,

$$\frac{\partial \Lambda}{\partial \mathbf{a}} = \mathbf{N} - \mathbf{M} \cdot \mathbf{a} = \mathbf{0}, \quad (19)$$

for which the solution is $\mathbf{a} = \mathbf{M}^{-1} \cdot \mathbf{N}$. Then the maximum of $\Lambda(\mathbf{r})$ over the extrinsic parameters is

$$\max_{\lambda_e} \{\Lambda\} = \frac{1}{2} \mathbf{N} \cdot \mathbf{M}^{-1} \cdot \mathbf{N}. \quad (20)$$

Having solved for \mathbf{a} , the extrinsic parameters can be explicitly derived from Eq. 6,

$$\zeta = \frac{1}{2} \left((a_1^2 + a_3^2)^{1/2} + (a_1^2 - a_2^2 + a_3^2 - a_4^2)^{1/2} \right), \quad (21a)$$

$$\psi = \frac{1}{2} \arctan \left(-a_3/a_1 \right), \quad (21b)$$

$$\iota = \arccos \left(-a_2/(2\zeta \sin \psi) \right), \quad (21c)$$

where $\zeta > 0$, $\psi \in [0, \pi]$, and $\iota \in [0, \pi]$.

3.2. Karush-Kuhn-Tucker conditions

The unconstrained maximum presented above is not the correct solution because there are indeed additional constraints on \mathbf{a} . From Eq. 6 we notice the fact that $\tan 2\psi = -a_3/a_1 = a_2/a_4$, therefore we have one nonlinear equality constraint (NEC),

$$a_1 a_2 + a_3 a_4 = 0. \quad (22)$$

It turns out that this condition also guarantees that the absolute value of the argument of the arccos function in Eq. 21 is not greater than one. Furthermore, to have a meaningful (square root of a nonnegative number) solution of ζ in Eq. 21 one additional nonlinear inequality constraint (NIEC)

$$a_1^2 - a_2^2 + a_3^2 - a_4^2 \geq 0, \quad (23)$$

is required. (Notice the arctan function is always meaningful.)

The *Karush-Kuhn-Tucker* (KKT) conditions provide a formal framework for solving constrained optimization problems that include equality and inequality constraints. Essentially, these conditions state that the solution is guaranteed to lie in the region where an inequality constraint is satisfied or on the boundary of this region. The boundary is obtained by turning the inequality constraint into an equality constraint.

Following the KKT prescription, the strategy for finding the solution of our problem is composed of two steps, as described below. The first step finds a solution that satisfies only the NEC. The solution to the second steps satisfy both NEC and NIEC. Only if the solution from the first step does not satisfy NIEC, is the second step taken.

Step 1: taking NEC into account. We can include the nonlinear equality constraint in the Lagrangian method as follows

$$\Lambda = \mathbf{a}^T \mathbf{N} - \frac{1}{2} \mathbf{a}^T \mathbf{M} \mathbf{a} + \frac{\mu}{2} \mathbf{a}^T \mathbf{D} \mathbf{a}, \quad (24)$$

where μ is the Lagrangian multiplier for the nonlinear constraint, and

$$\mathbf{D} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}. \quad (25)$$

Then, differentiating Eq. 24 with respect to \mathbf{a}

$$\frac{\partial \Lambda}{\partial \mathbf{a}} = \mathbf{N} - \mathbf{M} \cdot \mathbf{a} + \mu \mathbf{D} \cdot \mathbf{a} = \mathbf{0}, \quad (26)$$

and the multiplier μ

$$\frac{\partial \Lambda}{\partial \mu} = \frac{1}{2} \mathbf{a} \cdot \mathbf{D} \cdot \mathbf{a} = 0, \quad (27)$$

and solving Eq. 26 and Eq. 27, we get

$$\mathbf{a}_{\text{NEC}} = (\mathbf{M} - \mu \mathbf{D})^{-1} \mathbf{N}, \quad (28)$$

where μ can be solved by using one dimensional numerical root finding on

$$\mathbf{N}^T (\mathbf{M} - \mu \mathbf{D})^{-1} \mathbf{D} (\mathbf{M} - \mu \mathbf{D})^{-1} \mathbf{N} = 0. \quad (29)$$

Having obtained μ , we insert it back into Eq. 28 to get \mathbf{a} and then Λ from Eq. 17. We then check if this solution of \mathbf{a} violates the nonlinear inequality condition in Eq. 23. If not, it is already the correct solution; if yes, we then need to incorporate this quadratic inequality in the Lagrangian multiplier equation.

Step 2: taking NEC and NIEC into account. Since the solution of \mathbf{a} always appears at the boundary of the region satisfying the inequality, this constraint reduces to $a_1^2 - a_2^2 + a_3^2 - a_4^2 = 0$. As above we need to solve the following problem

$$\Lambda = \mathbf{a}^T \mathbf{N} - \frac{1}{2} \mathbf{a}^T \mathbf{M} \mathbf{a} + \frac{\mu}{2} \mathbf{a}^T \mathbf{D} \mathbf{a} + \nu \mathbf{a}^T \mathbf{B} \mathbf{a}, \quad (30)$$

where ν is the Lagrangian multiplier that takes care of the second quadratic condition, and

$$\mathbf{B} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}. \quad (31)$$

As before, taking the partial derivatives of Λ with respect to \mathbf{a} , μ , and ν will get

$$\mathbf{a}_{\text{NEC+NIEC}} = (\mathbf{M} - \mu\mathbf{D} - 2\nu\mathbf{B})^{-1}\mathbf{N}. \quad (32)$$

Here multiplier μ and ν can be solved simultaneously from the following two equations

$$\mathbf{N}^T(\mathbf{M} - \mu\mathbf{D} - 2\nu\mathbf{B})^{-1}\mathbf{D}(\mathbf{M} - \mu\mathbf{D} - 2\nu\mathbf{B})^{-1}\mathbf{N} = 0, \quad (33)$$

$$\mathbf{N}^T(\mathbf{M} - \mu\mathbf{D} - 2\nu\mathbf{B})^{-1}\mathbf{B}(\mathbf{M} - \mu\mathbf{D} - 2\nu\mathbf{B})^{-1}\mathbf{N} = 0. \quad (34)$$

To solve the above two equations simultaneously one needs a two-dimensional numerical root finding method. This step is, therefore, computationally more demanding than the one-dimensional method in step 1. Fortunately the chance of facing this case is very small in practice, and our strategy to handle it is to numerically solve the quadratically-constrained quadratic program (QCQP) with the active-set algorithm (Nocedal & Wright 2006) or the interior-point algorithm (Byrd et al. 1999). We choose the former in our demonstration in Sec. 5 for its computational efficiency, although the latter could be more accurate if the gradients and the Hessians of Λ and the constraints are provided. Either algorithm needs a starting point for the search, and the \mathbf{a}_{NEC} in step 1 is proved to be a good choice.

4. Particle Swarm Optimization (PSO)

In Sec. 3.2, we showed how to maximize the extrinsic parameters λ_e for a given set of intrinsic parameters λ_i . Maximization over λ_i requires a search over a $4 + N_p$ dimensional space. Due to the presence of noise and degeneracies in the intrinsic signal parameter space, $\Lambda(\mathbf{r})$ is a highly multi-modal function having a forest of local optima. Deterministic local optimization methods cannot be used to locate the global optimum of such a function. A brute force search using a grid of points is computationally prohibitive because the density of the grid must be high to tackle the large number of local optima while, at the same time, the number of grid points grows exponentially with the dimensionality of the search space.

The only feasible approach when it comes to multi-modal and high dimensional optimization problems is to use algorithms that employ some type of a stochastic search scheme.

A large class of these algorithms are modeled after biological systems (Engelbrecht 2005) with Particle Swarm Optimization (PSO) (Eberhart & Kennedy 1995) being among the popular choices. The use of PSO in gravitational wave data analysis was introduced in Wang & Mohanty (2010) where it was applied to compact binary inspiral searches for ground based detectors. In Mohanty (2012a,b), PSO was applied to the problem of GW burst detection where global optimization over a high dimensional search space is required. PSO has previously been used in pulsar timing analysis for the case of a two dimensional search space (Taylor et al. 2012; Lentati et al. 2013). Here, we use PSO in the context of the $(4 + N_p)$ dimensional search space. A brief description of PSO follows.

Consider the global maximization of a scalar function $f(x)$, called the *fitness* function, where $x \in S \subset \mathbb{R}^n$. S is called the *search space* which, for simplicity, will be assumed to be a hypercube: $S = [a, b] \otimes [a, b] \otimes \dots \otimes [a, b]$. In PSO, $f(x)$ is sampled at a fixed number of points and the coordinates of these points are evolved iteratively. Each point is called a *particle* and the set of particles is called a *swarm*. At each iteration step, indexed by an integer $k = 0, 1, \dots$, the fitness is evaluated at the current location of each particle.

Let $x_i(k)$, $i = 1, 2, \dots, N_{\text{part}}$, be the position of the i^{th} particle in a swarm of N_{part} particles at the iteration step k . The coordinates corresponding to $x_i(k)$ are $(x_{i,1}(k), \dots, x_{i,n}(k))$. The evolution equations for the swarm mimic in a simple way the behavior of real biological swarms (e.g., a flock of birds). Associated with each particle is a memory of the location where it found the best fitness value over its past history. This location, $p_i(k)$, is called *pbest* (“particle best”).

$$f(p_i(k)) = \min_{j=k, k-1, \dots, 0} f(x_i(j)) . \quad (35)$$

Associated with the swarm is a memory of the location where the swarm found the best fitness value over its past history. This location, $g(k)$, is called *gbest* (“global best”).

$$f(g(k)) = \min_{j=1, \dots, N_{\text{part}}} f(p_j(k)) . \quad (36)$$

Given $x_i(k)$, $p_i(k)$ and $g(k)$, the following equations are used to evolve the swarm.

$$x_i(k+1) = x_i(k) + v_i(k); \quad (37)$$

$$v_{i,j}(k+1) = \min(\max(y_{i,j}(k+1), -v_{\text{max}}), v_{\text{max}}) , \quad (38)$$

$$y_i(k+1) = w(k)v_i(k) + \mathbf{m}_{i,1}(p_i(k) - x_i(k)) + \mathbf{m}_{i,2}(g(k) - x_i(k)) , \quad (39)$$

where $v_i(k)$ is called the “velocity” of a particle. The second and third terms in Eq. 39, called “acceleration” terms, change the velocity in a random manner: $\mathbf{m}_{i,p}$, $p = 1, 2$, is a diagonal matrix, $\text{diag}(m_{p,i,1}, \dots, m_{p,i,n})$, such that $m_{p,i,k} \sim U[0, c_p]$ is drawn from a uniform

distribution over $[0, c_p]$, where c_p is a user specified parameter. The parameter $w(k)$ is called the “inertia” of a particle and it can change as the iteration progresses according to some specified law. At the termination of PSO, the highest fitness value found by the swarm, and the location of the particle with that fitness, make up the solution to the optimization problem.

The swarm is initialized by $x_{i,j}(0) = U[a, b]$ and $v_{i,j}(0) = U[-v'_{\max}, v'_{\max}]$. Usually, $v_{\max} = v'_{\max}$, but we fix them independently in this paper in order to promote greater exploration of S (see below). The use of the limiting speed v_{\max} prevents the swarm from exploding and leaving S . The same can be accomplished by a velocity constriction factor (Clerc & Kennedy 2002).

The physical meaning of the dynamical equations is fairly easy to grasp. In the absence of the acceleration terms, each particle simply moves in a straight line set by the vector $v_i(k=0)$. With the acceleration terms on, the particle is deflected on the average towards $pbest$ and $gbest$. Thus, each particle explores the search space under the competing pulls of moving independently of the swarm, which encourages exploration of the search space, and moving towards the best location found by the swarm, which encourages convergence (or *exploitation*).

In general, the behavior of the swarm transitions from exploration in the early phase to exploitation and convergence to an optimum in the late phase. A longer time spent exploring leads to a better chance of locating the global optimum but it also increases the computational cost of the method. Too soon a transition to exploitation may lead to premature convergence to a local optimum. The relative time spent in the two phases should be governed by the nature of the fitness function. However, since the degree of multi-modality is often unknown, it is best to err on the side of caution and extend the exploration phase.

One way to extend the exploration phase is to identify a set of neighbors for each particle and use the best value found within this neighborhood, called $lbest$ (“local best”), in Eq. 39 instead of $gbest$. By making each particle a member of multiple neighborhoods, information about $gbest$ leaks across the swarm but the rate at which this happens can be slowed down significantly compared to the case where each particle is constantly aware of $gbest$. There are many schemes for selecting neighborhoods in the PSO literature with the simplest being the *ring topology*: the particle indices are put on a circle and the neighborhood of each particle consists of $(m-1)/2$ particles on each side with m being the user specified size of each neighborhood.

The settings for most of the parameters of the PSO algorithm outlined above have been found to be quite robust across a wide range of fitness functions (Bratton & Kennedy

2007). This is, in fact, one of the attractive features of PSO as it considerably lessens the effort needed for tuning the algorithm. In the present paper we choose the following settings: $N_{\text{part}} = 40$, $c_1 = c_2 = 2.0$, $m = 3$, $v_{\text{max}} = (b - a)/5$, $v'_{\text{max}} = (b - a)/2$, $w(k) = 0.9 - 0.5(k/(N_{\text{iter}} - 1))$, where N_{iter} is the total number of iterations. In addition to fixing the PSO parameters, the behavior of particles crossing the boundary of S must be prescribed. We use the so-called “let them fly” boundary condition in which the fitness of the particle is simply set to $-\infty$ while it is outside S . This ensures that boundary crossing particles continue to behave according to the PSO dynamical equations while eventually being drawn back into S after a small number of iterations..

Although successful convergence to the global maximum is not guaranteed for any finite value for N_{iter} , increasing N_{iter} increases the probability of success. Thus, N_{iter} should be chosen to make the probability of success sufficiently close to unity while keeping computational costs within limits. This apparently straightforward task is complicated, however, by the fact that the value of the global maximum of Λ is not known *a priori* and, hence, we do not know when PSO has succeeded. This problem can be overcome to some extent for simulated data realizations following Mohanty (2012b). A PSO run is declared successful if it finds a value of Λ that is better than the one at the known location of the true signal in the search space. The underlying idea here is that such a condition should hold for any good parameter estimation algorithm since, otherwise, estimation would never incur an error due to noise. This criterion can be used to tune N_{iter} using simulated data realizations. One then hopes that the same setting would also lead to a large probability of success for real data provided that it is well modeled by the simulations. Following this criterion, we found that $N_{\text{iter}} = 2000$ was sufficient to give a fairly high probability of convergence for the simulations described below. As we discuss in Sec. 5.4, this is a reasonable but not a foolproof choice since a small fraction of cases do lead to failure according to this criterion. However, it is possible to overcome these failures by running PSO multiple times on the same data realization. This computationally expensive strategy is not followed for the bulk of our simulations in this paper although it is clearly an option that should be used in any analysis of real data.

5. Applications

We apply the algorithm described above to simulated data from a PTA using the parameters of eight pulsars from the NANOGrav catalog (Demorest et al. 2013). In all the cases considered below, the source is assumed to be a circular binary located at right ascension $\alpha = 1.985$ rad (7 hr 35 min) and declination $\delta = 0.625$ rad (35.83°), having an orbital angular frequency $\omega = 8.0$ rad yr $^{-1}$ ($\omega_{gw} = 16.0$ rad yr $^{-1}$), and initial phase $\varphi_0 = 1.6$ rad (91.67°)

at the start of observations. We set the inclination angle $\iota = 0^\circ$, leading to a circularly polarized GW tensor, and the polarization angle $\psi = 45^\circ$. The duration of the simulated observation is 5 years, with uniform biweekly cadence (number of samples $N_I = N_s = 128$ for each pulsar). The noise-free timing residual (the *signal*) induced by this GW source is calculated for each pulsar in the PTA following Eq. 5. Pseudo-random sequences of white Gaussian noise with a standard deviation $\sigma_n = 10^{-8}$ sec are added to the signal calculated for each pulsar to generate realizations of PTA data. To characterize the strength of the signal in the data, we use the network SNR of the signal defined as

$$\rho_n = \left(\sum_{I=1}^{N_p} \langle s^I | s^I \rangle_I \right)^{1/2} = \frac{1}{\sigma_n} \left(\sum_{I=1}^{N_p} \sum_{k=1}^{N_s} (s_k^I)^2 \right)^{1/2}. \quad (40)$$

For studying the statistical properties of the detection statistic and the estimated parameters, three different values of ρ_n are chosen and 500 independent realizations of data are generated for each value of ρ_n . Table 1 shows the values of ρ_n used, and corresponding to each, mean and standard deviation of the astrophysically interesting parameters (α , δ , ω_{gw}). It also lists the corresponding recovered SNR ρ_r calculated by replacing the s_k^I in Eq. 40 by the reconstructed signal. Further details about each scenario are reported separately in the following subsections.

Scenario	ρ_n	α		δ		ω_{gw}		ρ_r	
		mean	std.	mean	std.	mean	std.	mean	std.
strong	100	1.91	0.121	0.54	0.140	16.007	0.0088	100.1	1.05
moderate	10	1.85	0.522	0.32	0.460	15.999	0.0844	10.7	1.00
weak	5	2.15	1.004	0.20	0.672	15.184	0.1999	6.4	0.77

Table 1: The mean and the standard deviation of the three parameters α , δ and ω_{gw} calculated from 500 realizations of the three scenarios with different ρ_n , and the corresponding recovered SNR ρ_r .

In order to detect or set up upper limits, one needs to know the distribution of the detection statistic under \mathcal{H}_0 and \mathcal{H}_λ . This involves finding the distribution of the extremum (maximum) of the likelihood ratio Λ , which is, in general, difficult to obtain. In this work we choose to use the Monte-Carlo simulation with 500 realizations for the \mathcal{H}_λ mentioned above and 1000 realizations for \mathcal{H}_0 to obtain the respective distributions. The distribution for \mathcal{H}_0 can not be fitted by a known simple PDF, considering the distribution for the $\rho_n = 5$ case can be fitted by a Rician distribution $R(\nu, \sigma)$, with the noncentrality parameter $\nu = 20.1$ and the scale parameter $\sigma = 5.14$. It turns out that the distribution converges to a normal

distribution with the signal strength increases: for the $\rho_n = 10$ case, the distribution can be fitted by $\mathcal{N}(\mu = 57.5, \sigma = 10.7)$; for the $\rho_n = 100$ case, $\mathcal{N}(\mu = 5006.3, \sigma = 105.6)$.

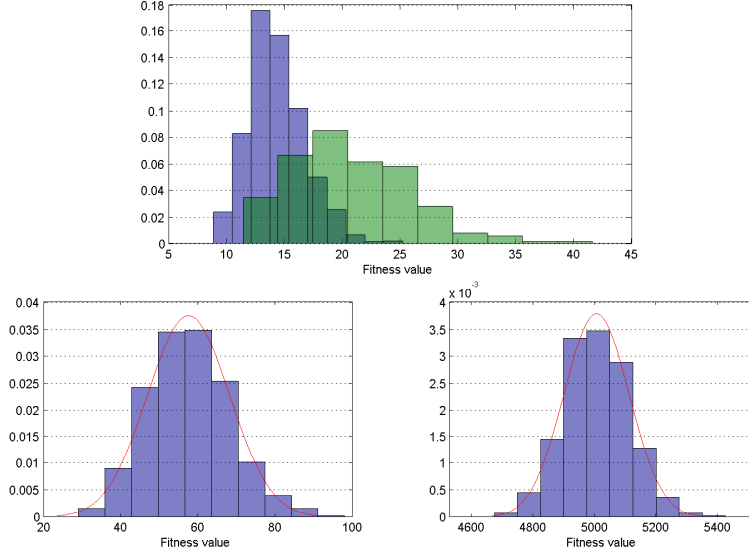


Fig. 2.— Histograms of the detection statistic. Blue histogram in upper panel is for \mathcal{H}_0 case, green histogram in the same panel is for $\rho_n = 5$ case. Histogram in lower left panel is for the $\rho_n = 10$ case. Histogram in lower right panel is for the $\rho_n = 100$ case. The red curves in the lower panels show the normal distribution with the best fit parameters mentioned in the text respectively.

5.1. Strong signal

The case $\rho_n = 100$ could arise, for example, from a SMBHB with the chirp mass $\mathcal{M}_c \approx 10^9 M_\odot$, at a distance $D \approx 10$ Mpc. The orbital period $P = 0.785$ yrs. Fig. 3 shows a typical realization of the simulated timing residuals for the eight pulsars (thick gray). As we can see, the magnitude and the phase of the noise-free timing residual (black dash) is different for the different pulsars in the PTA. The signal in most of the pulsars is comparable to or even stronger than the noise. To obtain the reconstructed signals (solid curve), we take the estimated intrinsic parameters and derive the extrinsic parameters by Eq. 21. Using the estimated intrinsic and extrinsic parameters, the estimated signal is found from Eq. 5. The estimated signal agrees with the injected signal very well for most pulsars, except for PSR J1909-3744, which has a signal amplitude about two orders of magnitude smaller than the others. This pulsar is almost (oppositely) aligned with the GW source (separation angle $\theta = 174^\circ$), therefore it is insensitive to this source and has insignificant contribution to the

MLR statistic. The mean of the recovered SNR in Table 1 is larger than the network SNR means that the detection statistic performs well in recovering the signals. Fig. 2 shows the distribution of the detection statistic values for the signal present and absent cases. From the large separation of these distributions, it is clear that the detection probability Q_d is nearly unity if the threshold is chosen to be the highest value found for the signal absent case. (The false alarm probability for this choice of threshold is $\simeq 10^{-3}$.)

Fig. 4 shows the distributions of the estimated intrinsic parameters $\{\alpha, \delta, \omega_{gw}, \varphi_0, \varphi_I\}$ ($I = 1, 2, \dots, N_p = 8$) from the 500 realizations. The red vertical line marks the true value of the parameter used in the simulations. The dash vertical line marks the mean value. It appears that the right ascension, the declination and the frequency of GW can be accurately estimated with the standard deviations $\sigma_\alpha \approx 7^\circ$, $\sigma_\delta \approx 8^\circ$, and $\sigma_{\omega_{gw}} < 0.01 \text{ rad yr}^{-1}$. The one sigma contour of the source location distribution encloses an area of $\sim 40 \text{ deg}^2$ on the sky. In contrast the initial orbital phase and the pulsar phases are poorly estimated even for this strong signal scenario. Therefore, the pulsar phase parameters are truly nuisance parameters that should either be maximized or marginalized over when constructing detection algorithms. It is not practical to estimate the pulsar distance from the analysis of continuous gravitational waves. For PSR J1909-3744, the distribution of the pulsar phase is less concentrated than the other pulsars due to the reason mentioned above. It is important to note that the initial orbital phase φ_0 and the pulsar phases φ_I are directional (circular) variables, and their distributions should be wrapped around at the end points 0 and π . For example, in the histogram of the pulsar phase for PSR J0613-0200, the counts for values of φ_I below 1.5 radians should be appended after the histogram ending at $\varphi_I = \pi$. Although right ascension and declination are also circular variables, the distributions are narrow and centered on values not close to the end points of the range $[0, \pi]$, therefore the wrapping of the histograms can be ignored.

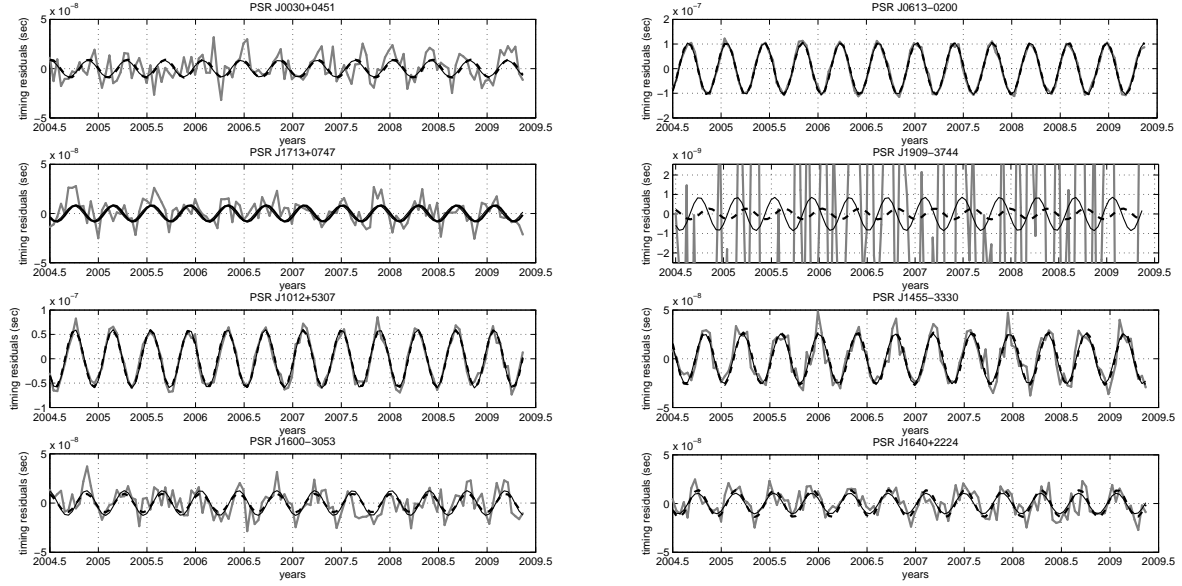


Fig. 3.— A data realization showing the simulated timing residuals (thick gray) and signal (dash black) for all pulsars. The network signal to noise ratio is $\rho_n = 100$. The reconstructed signals are shown as solid curves. For most pulsars, the true and reconstructed signal are almost indistinguishable from each other.

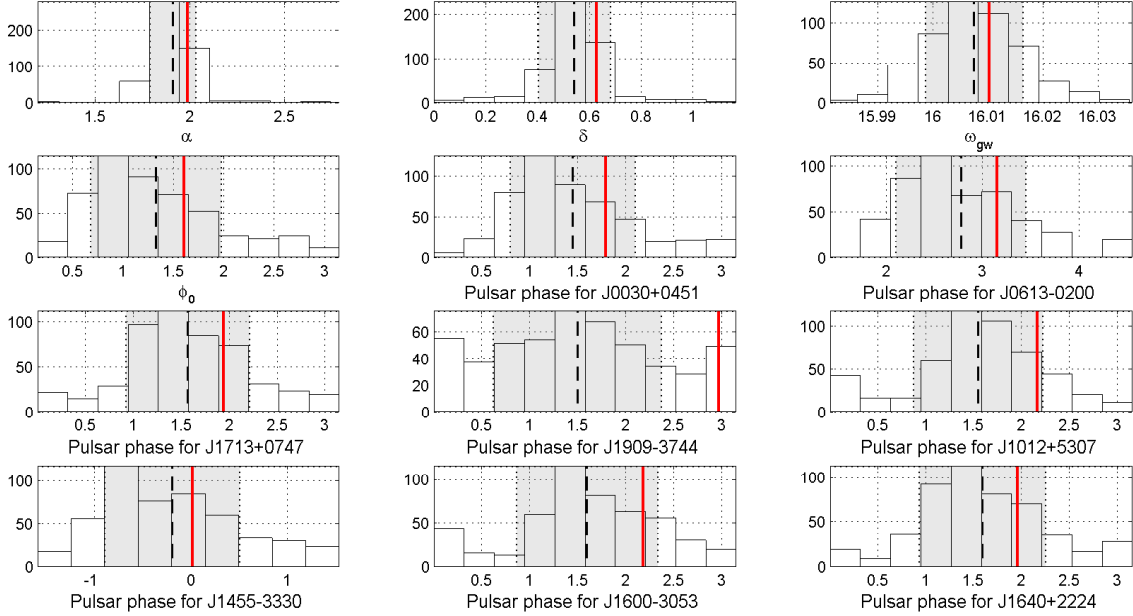


Fig. 4.— Histograms of the estimated intrinsic parameters for the case with the network signal to noise ratio $\rho_n = 100$. The red vertical line marks the true value of the parameter used in the simulations, the dash vertical line marks the mean value, and shaded area covers the one sigma uncertainty. The total number of trials is 500.

5.2. Moderate signal

In this scenario, the network signal to noise ratio $\rho_n = 10$. This corresponds, for example, to a SMBHB with chirp mass $\mathcal{M}_c \approx 10^9 M_\odot$ at a distance $D \approx 100$ Mpc. Fig. 5 shows a realization of the simulated timing residuals for this scenario. The recovered signal agrees well with the injected signal in 5 of the pulsars that have the strongest signals, comparing the others having various levels of offsets in phase and changes in amplitude. Although the estimated and true signals now agree well in only about 5 of the pulsars, the recovered SNR (c.f., Table 1) is still high. This indicates that a strong detection is still possible in this scenario, but estimation accuracy has worsened. Fig. 6 shows the distributions of the estimated intrinsic parameters. The estimation of the sky location is considerable worse than the $\rho_n = 100$ scenario, with the standard deviations $\sigma_\alpha \approx 30^\circ$, $\sigma_\delta \approx 26^\circ$ respectively. The one sigma contour of the source location distribution encloses an area of $\sim 800 \text{ deg}^2$ on the sky. However the GW frequency can still be estimated quite accurately with $\sigma_{\omega_{gw}} < 0.1 \text{ rad yr}^{-1}$.

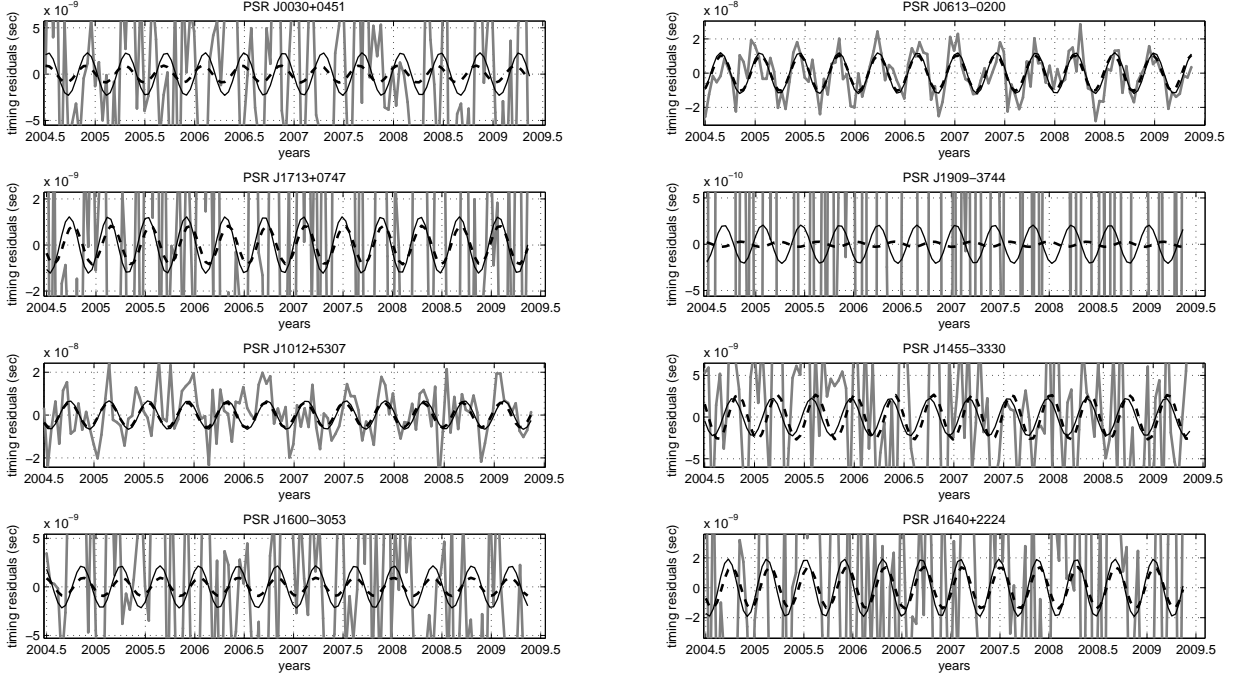


Fig. 5.— A data realization showing the simulated timing residuals (thick gray) and signal (dash black) for all pulsars. The network signal to noise ratio is $\rho_n = 10$. The reconstructed signals are shown as solid curves. For some pulsars, such as PSR J0030+0451 and J1909-3744, we have zoomed into the noise in the subplots, so that the signal can be manifested.

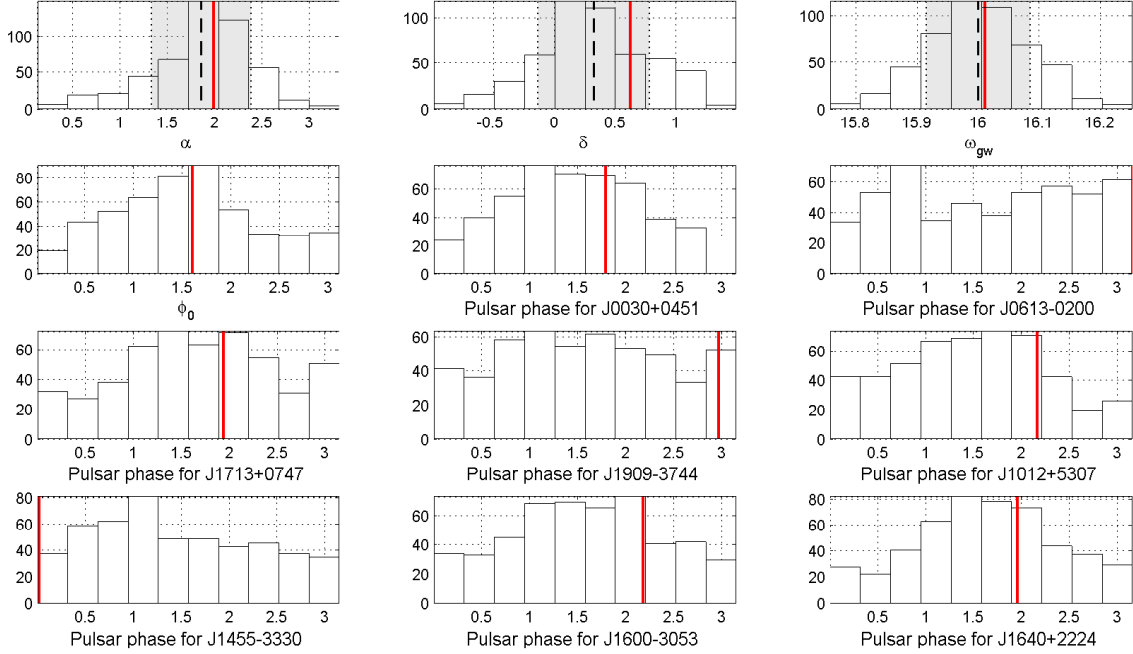


Fig. 6.— Histograms of the estimated intrinsic parameters for the case with the network signal to noise ratio $\rho_n = 10$. The red vertical line marks the true value of the parameter used in the simulations. The total number of trials is 500.

5.3. Weak signal

In this scenario, the signal strength is weak with a network signal to noise ratio $\rho_n = 5$. This corresponds, for example, to a SMBHB with chirp mass $\mathcal{M}_c \approx 10^9 M_\odot$ and the distance $D \approx 10^3$ Mpc. Fig. 7 shows a realization of the simulated timing residuals in which the noise is seen to dominate the signal for all the pulsars. This realization illustrates the most likely case for the current level of sensitivity for gravitational wave detection using pulsar timing arrays. Fig. 8 shows the distribution of the estimated intrinsic parameters. Consequently, the reconstructed signals for only two pulsars agree reasonably well with the true signals, whereas for most of the pulsars the match is quite poor. From Fig. 2, the detection probability $Q_d \simeq 18.6\%$ if we choose the threshold to be the highest value obtained for the noise-only case. The one sigma contour encloses an area more than 10^3 deg^2 , a large fraction of the sky. The estimation of GW frequency is tolerable. Therefore in this case, although the signal is still strong enough to be detected, it is not strong enough to be localized.

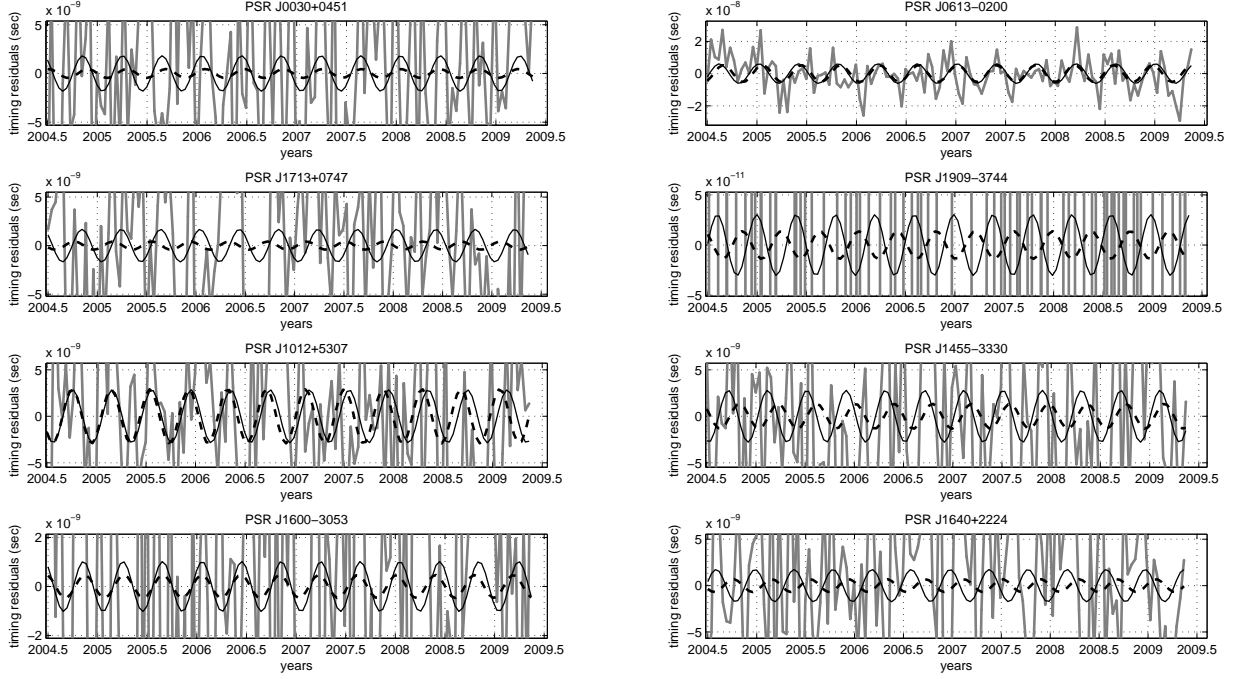


Fig. 7.— A data realization showing the simulated timing residuals (thick gray) and signal (dash black) for all pulsars. The network signal to noise ratio is $\rho_n = 5$. The reconstructed signals are shown as solid curves. For all pulsars, we have zoomed into the noise in the subplots, so that the signal can be manifested.

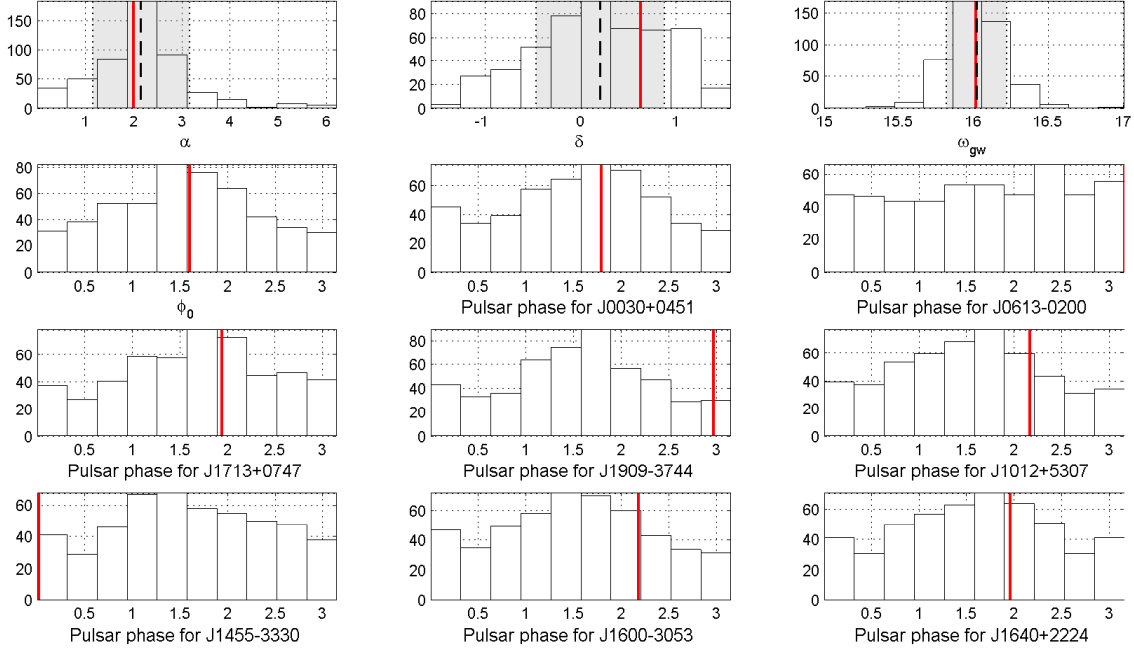


Fig. 8.— Histograms of the estimated intrinsic parameters for the case with the network signal to noise ratio $\rho_n = 5$. The red vertical line marks the true value of the parameter used in the simulations. The total number of trials is 500.

5.4. Improvements from multiple runs of PSO

As mentioned in Sec. 4, we tuned the number of iterations (N_{iter}) in PSO such that it converges to the global maximum for the bulk of the trials in our simulations. The evidence for a successful convergence is the achievement of a better fitness value than the one at the known location of the true signal. Fig. 9 shows the fitness value calculated with the estimated parameters (MLR) using the algorithm and PSO versus the fitness value calculated with the true parameters (LR) used for the 500 realizations of the simulated data. The blue cross above the red diagonal line means that the MLR is larger than the LR for the realization. As we can see, this is true for all realizations for $\rho_n = 10$ and $\rho_n = 5$ cases, comparing about 20% of the realizations for $\rho_n = 100$ are below the red line.

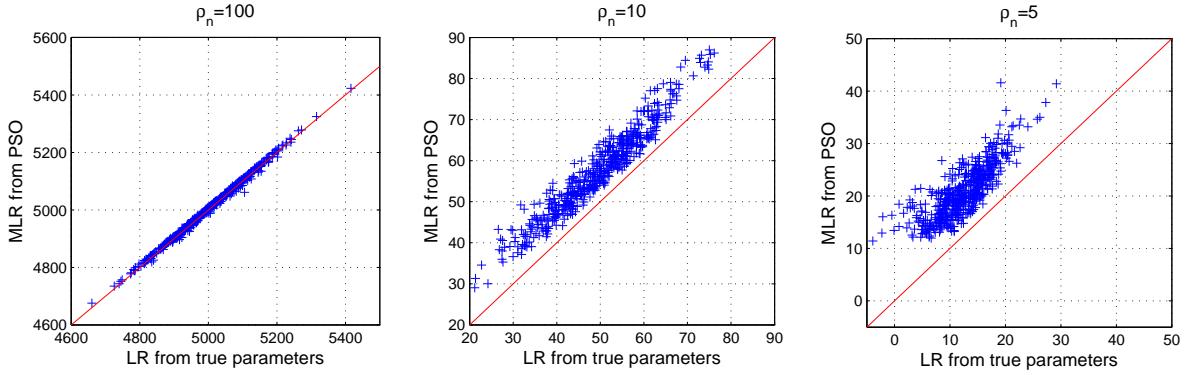


Fig. 9.— MLR calculated with estimated parameters versus LR calculated with true parameters for the 500 realizations for the $\rho_n = 100$, 10 and 5 scenarios respectively.

It is possible to remedy the small fraction of failures seen in Fig. 9 by a simple strategy: perform multiple independent runs of PSO on the same data realization and use the output from the run that finds the best fitness value. Although it makes perfect sense to use this strategy in any implementation involving real data, it is computationally too expensive for a simulation involving hundreds of data realizations. Here we only demonstrate the viability of this strategy by applying it to a small fraction of the simulated data realizations.

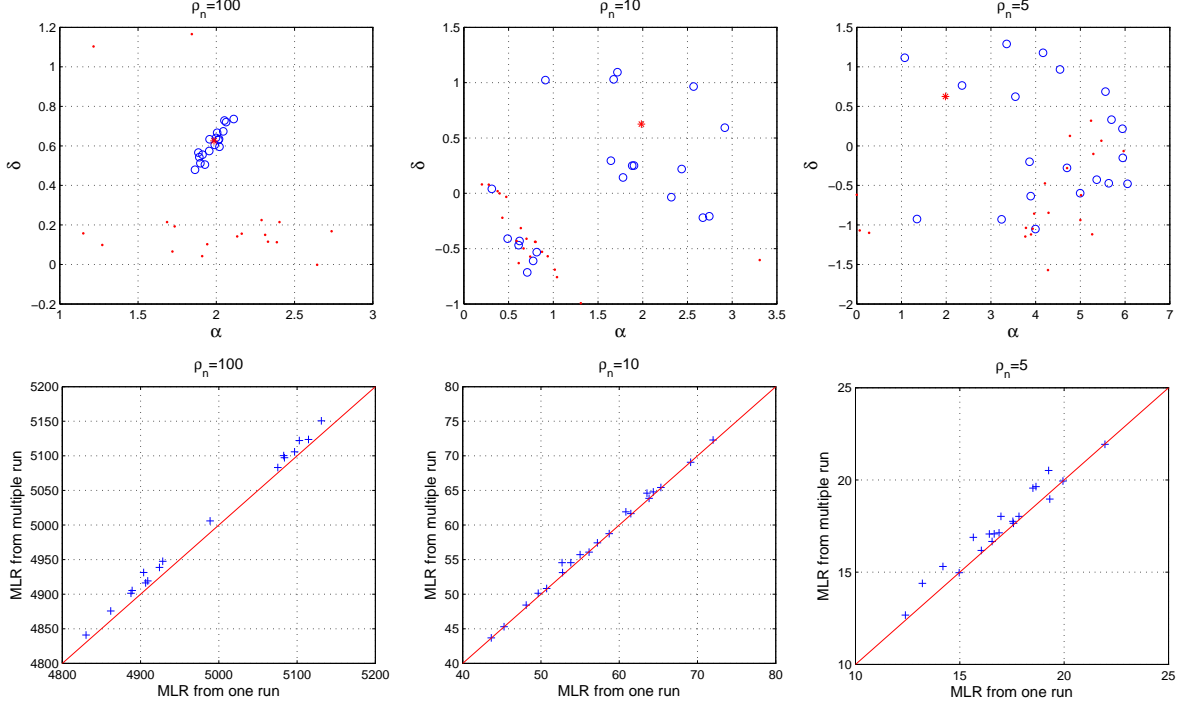


Fig. 10.— In each column, the top subpanel shows estimated sky locations, with results from the rerun of the method (10 independent PSO runs) shown as circles and the results from the earlier runs (1 PSO run) shown as dots. The X-axis shows right ascension and Y-axis shows declination. The true location is shown as an asterisk. The bottom panel in each column shows the MLR from this method versus the MLR from the earlier runs.

For each scenario in Table 1, we select 4% of the realizations for which the estimated sky locations departed most from the true location. We then rerun the detection and estimation method, with 10 independent runs of PSO instead of one, on each of these data realizations, and use the results from the run that terminates with the best fitness value. Fig. 10 compares the estimated sky locations and fitness values found by the old and new runs. For the $\rho_n = 100$ case, the estimated sky locations clearly improve significantly. For the $\rho_n = 10$ and $\rho_n = 5$ cases, the improvement is not as dramatic. However, we see that the new estimates are dispersed around the true location in a much more symmetrical way than the previous ones. Although this does not improve the variance of the estimates, it clearly reduces bias. For all values of ρ_n , the MLR is improved for most data realizations. Note that the new set of runs of PSO are independent of the old ones as they start off with different initial conditions. Hence, it is not surprising that the new strategy may perform marginally worse in some rare cases. Again, increasing the number of independent PSO runs a bit more should

considerably diminish the chances of this happening. We also see that the typical increase of MLR is smaller than the bin size of the histograms in Fig. 2. Therefore the distribution of the detection statistic will not change considerably if we use multiple runs of PSO for all the realizations. Estimation performance of the method, however, should improve somewhat. (The significant improvement for the $\rho_n = 100$ case is tempered by the fact that it represents an extremely unlikely signal strength.)

6. Conclusions

We have presented a coherent network analysis method based on the GLRT for the detection and estimation of monochromatic continuous gravitational waves using a pulsar timing array. The method explicitly includes pulsar phase parameters in the maximization of the likelihood ratio. The resulting high dimensional optimization problem is successfully addressed using Particle Swarm Optimization. By keeping the pulsar phases as intrinsic parameters in the maximization, the method resembles the \mathcal{F} -statistic in the choice of extrinsic parameters. However, our results show that the pulsar phases are uninformative and should be treated as nuisance parameters in future studies. Maximization or marginalization over these parameters will result in a detection statistic that is radically different from the \mathcal{F} -statistic. Quantifying the performance of this statistic will be the subject of future work.

Even if one treats the pulsar phases as extrinsic parameters, the number of intrinsic parameters will not be small. Hence, stochastic optimization methods such as PSO will still be required. We have shown that PSO already works well with only minimal effort required in its tuning. Therefore, we are confident that it will continue to be useful when we shift to a method that treats pulsar phases as extrinsic parameters.

A limitation of the present study is that the simulated observations are evenly sampled with biweekly cadence, and have stationary white Gaussian noise with zero mean and the same variance for all pulsars. However, since the method presented here works entirely in the time domain, no modifications are needed to switch to irregularly sampled data. Given that the correct way to proceed appears to involve changing the pulsar phases into extrinsic parameters, simulations with more realistic data models are best postponed until such a method is realized. Finn (2001) has shown already that coherent methods, such as the one presented here, are robust against non-Gaussianity in timing residual noise. Thus, the Gaussian noise assumption in our simulations, although incommensurate with expectations about real data (Wang & et al. 2014), is not a serious flaw.

So far, we have assumed that the timing residuals are a superposition of the signal

and the noise. In reality, the timing residuals are obtained by a weighted least square fit as mentioned in Sec. 1. Due to the covariance of the GW signal with some of the fitting parameters, the power of GW signal may be absorbed in the fitting model, and the signal in the residuals may be changed. Besides, the fitting procedure can also change the statistics (e.g. covariance matrix) of the noise, especially the noise component can become non-stationary in the timing residual even though the noise in the TOA is stationary. In order to detect gravitational waves, one must take account of these fitting effects. One approach is to use the projection operator (matrix) \mathbf{R} define in Demorest et al. (2013). It only depends on the fitting model and the weighting matrix, not the specific value of the data. This allows us to study the fitting effects by simulations. These factors should be taken into account in the analysis when we apply this algorithm to the real data.

Another element required in real data analysis but absent from our study is noise characterization. In this paper, we assumed that the noise parameters are known *a priori* or can be fixed at their independently estimated values. A more comprehensive approach would be to include them as part of the estimation process along with the signal parameters. Though straightforward in terms of the formalism, this will increase the dimensionality of the search space for PSO significantly (e.g., from 12 to 52 for a PTA of 8 pulsars). PSO is routinely used for optimization problems with comparable dimensionality, but it remains to be seen if the standard variant of PSO used in this paper will continue to work successfully when the above extension is made.

The method we have presented does not account for elliptical orbits or evolution of the orbital parameters. It is expected that various dissipation effects (including GWs) would have circularized the orbit by the time it enters the sensitive frequency band of PTA. The evolution of orbital parameters becomes a progressively important consideration as the signal frequency, linked to the orbital frequency, increases. The effect of orbital evolution during the travel time of radio pulses to Earth is that the pulsar term corresponds to a lower orbital frequency (earlier stage of binary source) than the one in the Earth term. In addition, according to Eq. 7, the amplitude of the pulsar term will be larger than the Earth term. These considerations can be accomodated in the analysis by the introduction of one extra intrinsic parameters, which will not affect the performance of PSO. The estimation errors of all the parameters, however, will worsen as a result. This work is subject to our future investigations.

7. Acknowledgments

This work was supported by the National Science Foundation under PIRE grand 0968296. The contribution of SDM to this paper is supported by NSF awards PHY-1205585 and HRD-0734800. We are grateful to the members in the NANOGrav for helpful comments and discussions.

REFERENCES

- Abadie, J., Abbott, B. P., Abbott, R., Abbott, T. D., Abernathy, M., Accadia, T., Acernese, F., Adams, C., Adhikari, R., Affeldt, C., & et al. 2012a, Phys. Rev. D, 85, 122007
- . 2012b, Phys. Rev. D, 85, 082002
- Abbott, B. P., Abbott, R., Adhikari, R., Ajith, P., Allen, B., Allen, G., Amin, R. S., Anderson, S. B., Anderson, W. G., Arain, M. A., & et al. 2009, Reports on Progress in Physics, 72, 076901
- Accadia, T., Acernese, F., Antonucci, F., & et al. 2011, Classical and Quantum Gravity, 28, 114002
- Arzoumanian, Z., Brazier, A., Burke-Spolaor, S., Chamberlin, S. J., Chatterjee, S., Cordes, J. M., Demorest, P. B., Deng, X., Dolch, T., Ellis, J. A., Ferdman, R. D., Finn, L. S., Garver-Daniels, N., Jenet, F., Jones, G., Kaspi, V. M., Koop, M., Lam, M., Lazio, T. J. W., Lommen, A. N., Lorimer, D. R., Luo, J., Lynch, R. S., Madison, D. R., McLaughlin, M., McWilliams, S. T., Nice, D. J., Palliyaguru, N., Pennucci, T. T., Ransom, S. M., Sesana, A., Siemens, X., Stairs, I. H., Stinebring, D. R., Stovall, K., Swiggum, J., Vallisneri, M., van Haasteren, R., Wang, Y., & Zhu, W. W. 2014, ArXiv e-prints
- Babak, S., & Sesana, A. 2012, Phys. Rev. D, 85, 044034
- Bratton, D., & Kennedy, J. 2007, in Swarm Intelligence Symposium, 2007. SIS 2007. IEEE, IEEE, 120–127
- Byrd, R. H., Hribar, M. E., & Nocedal, J. 1999, SIAM Journal on Optimization, 9, 877
- Clerc, M., & Kennedy, J. 2002, Evolutionary Computation, IEEE Transactions on, 6, 58
- Corbin, V., & Cornish, N. J. 2010, ArXiv e-prints

- Cutler, C., & Schutz, B. F. 2005, *Phys. Rev. D*, 72, 063006
- Degallaix, J., Accadia, T., Acernese, F., Agathos, M., Allocca, A., & et al. 2013, in *Astronomical Society of the Pacific Conference Series*, Vol. 467, 9th LISA Symposium, ed. G. Auger, P. Binétruy, & E. Plagnol, 151
- Demorest, P. B., Ferdman, R. D., Gonzalez, M. E., Nice, D., Ransom, S., Stairs, I. H., Arzoumanian, Z., Brazier, A., Burke-Spolaor, S., Chamberlin, S. J., Cordes, J. M., Ellis, J., Finn, L. S., Freire, P., Giampanis, S., Jenet, F., Kaspi, V. M., Lazio, J., Lommen, A. N., McLaughlin, M., Palliyaguru, N., Perrodin, D., Shannon, R. M., Siemens, X., Stinebring, D., Swiggum, J., & Zhu, W. W. 2013, *ApJ*, 762, 94
- Eberhart, R., & Kennedy, J. 1995, in *Micro Machine and Human Science, 1995. MHS'95., Proceedings of the Sixth International Symposium on*, IEEE, 39–43
- Edwards, R. T., Hobbs, G. B., & Manchester, R. N. 2006, *MNRAS*, 372, 1549
- Ellis, J. A. 2013, *Classical and Quantum Gravity*, 30, 224004
- Ellis, J. A., Siemens, X., & Creighton, J. D. E. 2012, *ApJ*, 756, 175
- Engelbrecht, A. P. 2005, *Fundamentals of computational swarm intelligence*, Vol. 1 (Wiley Chichester)
- Ferdman, R. D., van Haasteren, R., Bassa, C. G., Burgay, M., Cognard, I., Corongiu, A., D’Amico, N., Desvignes, G., Hessels, J. W. T., Janssen, G. H., Jessner, A., Jordan, C., Karuppusamy, R., Keane, E. F., Kramer, M., Lazaridis, K., Levin, Y., Lyne, A. G., Pilia, M., Possenti, A., Purver, M., Stappers, B., Sanidas, S., Smits, R., & Theureau, G. 2010, *Classical and Quantum Gravity*, 27, 084014
- Finn, L. S. 2001, *Phys. Rev. D*, 63, 102001
- Foster, R. S., & Backer, D. C. 1990, *ApJ*, 361, 300
- Gregory, P. 2010, *Bayesian Logical Data Analysis for the Physical Sciences*
- Hellings, R. W., & Downs, G. S. 1983, *ApJ*, 265, L39
- Jaranowski, P., Królak, A., & Schutz, B. F. 1998, *Phys. Rev. D*, 58, 063001
- Jenet, F. A., Creighton, T., & Lommen, A. 2005a, *ApJ*, 627, L125
- Jenet, F. A., Hobbs, G. B., Lee, K. J., & Manchester, R. N. 2005b, *ApJ*, 625, L123

- Jenet, F. A., Lommen, A., Larson, S. L., & Wen, L. 2004, *ApJ*, 606, 799
- Kay, S. 1998, *Fundamentals of Statistical Signal Processing, Volume II: Detection Theory* (Prentice Hall)
- Lentati, L., Alexander, P., Hobson, M. P., Taylor, S., Gair, J., Balan, S. T., & van Haasteren, R. 2013, *Phys. Rev. D*, 87, 104021
- Lorimer, D. R., & Kramer, M. 2004, *Handbook of Pulsar Astronomy*
- Maggiore, M. 2007, *Gravitational Waves: Volume 1: Theory and Experiments* (Oxford University Press)
- Manchester, R. N. 2013, *ArXiv e-prints*
- Manchester, R. N., Hobbs, G., Bailes, M., Coles, W. A., van Straten, W., Keith, M. J., Shannon, R. M., Bhat, N. D. R., Brown, A., Burke-Spolaor, S. G., Champion, D. J., Chaudhary, A., Edwards, R. T., Hampson, G., Hotan, A. W., Jameson, A., Jenet, F. A., Kesteven, M. J., Khoo, J., Kocz, J., Maciesiak, K., Osłowski, S., Ravi, V., Reynolds, J. R., Sarkissian, J. M., Verbiest, J. P. W., Wen, Z. L., Wilson, W. E., Yardley, D., Yan, W. M., & You, X. P. 2013, *PASA*, 30, 17
- Mohanty, S. D. 2012a, *The Astronomical Review*, 7, 020000
- . 2012b, *The Astronomical Review*, 7, 040000
- Nocedal, J., & Wright, S. J. 2006, *Numerical Optimization*
- Sathyaprakash, B., & Schutz, B. F. 2009, *Living Reviews in Relativity*, 12
- Seoane, P. A., Aoudia, S., Audley, H., Auger, G., & et al. 2013, *ArXiv e-prints*
- Sesana, A., & Vecchio, A. 2010, *Phys. Rev. D*, 81, 104008
- Sesana, A., Vecchio, A., & Volonteri, M. 2009, *MNRAS*, 394, 2255
- Somiya, K. 2012, *Classical and Quantum Gravity*, 29, 124007
- Springel, V., White, S. D. M., Jenkins, A., Frenk, C. S., Yoshida, N., Gao, L., Navarro, J., Thacker, R., Croton, D., Helly, J., Peacock, J. A., Cole, S., Thomas, P., Couchman, H., Evrard, A., Colberg, J., & Pearce, F. 2005, *Nature*, 435, 629
- Sudou, H., Iguchi, S., Murata, Y., & Taniguchi, Y. 2003, *Science*, 300, 1263
- Taylor, J. H., & Weisberg, J. M. 1982, *ApJ*, 253, 908

—. 1989, *ApJ*, 345, 434

Taylor, S. R., Gair, J. R., & Lentati, L. 2012, ArXiv e-prints

Verbiest, J. P. W., Bailes, M., Coles, W. A., Hobbs, G. B., van Straten, W., Champion, D. J., Jenet, F. A., Manchester, R. N., Bhat, N. D. R., Sarkissian, J. M., Yardley, D., Burke-Spolaor, S., Hotan, A. W., & You, X. P. 2009, *MNRAS*, 400, 951

Waldman, S. J. 2011, ArXiv e-prints

Wang, Y., & et al. 2014, in preparation

Wang, Y., & Mohanty, S. D. 2010, *Phys. Rev. D*, 81, 063002

Yardley, D. R. B., Hobbs, G. B., Jenet, F. A., Verbiest, J. P. W., Wen, Z. L., Manchester, R. N., Coles, W. A., van Straten, W., Bailes, M., Bhat, N. D. R., Burke-Spolaor, S., Champion, D. J., Hotan, A. W., & Sarkissian, J. M. 2010, *MNRAS*, 407, 669

Zhu, W. W., & et al. 2014, to be submitted